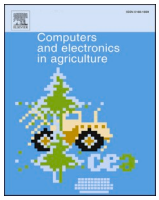
Computers and Electronics in Agriculture 205 (2023) 107633

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag



Remote sensing crop group-specific indicators to support regional yield forecasting in Europe

Giulia Ronchetti [a,*], Giacinto Manfron [a,*], Christof J. Weissteiner [b], Lorenzo Seguini [a], Luigi Nisini Scacchiafichi [a], Lorenzo Panarello [a], Bettina Baruth [a]

[a] European Commission, Joint Research Centre, 21027 Ispra, VA, Italy
[b] European Commission, European Research Executive Agency, 1049 Brussels, Belgium

ARTICLE INFO

Keywords:
Crop masks
Crop yield forecasting
Crop monitoring
Regional forecast
Remote Sensing
NDVI

ABSTRACT

Operational crop yield forecasting services typically provides crop yield forecasts based on regression models between official yields and agro-environmental variables, among which meteorological data, crop simulation model or satellite-derived indicators. The reliability with which these variables infer on yields depends, among many other factors, also on their aggregation in the space domain, for example on the type of the crop masks utilized in the aggregation process from point to regional scale. This work investigates how the yield explanatory power of satellite-derived indicators is changing moving from time-stable arable land masks to annual crop group-specific masks. We compare in particular the linkage between time series of regional crop yield in Europe and the Normalized Difference Vegetation Index (NDVI) from Moderate Resolution Imaging Spectroradiometer (MODIS), when generic arable land masks or crop group-specific and year-specific information are applied to aggregate pixel values at regional level. Regional (Eurostat level NUTS-2) crop yield statistics were collected from official databases for the period 2003–2019, while NDVI data were derived from MODIS daily products at 250 m spatial resolution for the same reference period. Regional NDVI profiles were retrieved by averaging single pixels time series according to the information of five crop masks, including generic arable land masks, crop group-specific static masks (separating winter and spring crops from summer crops), and annual crop group-specific masks (distinguishing between the two crop groups and varying in time). A compared correlation analysis between yield data and regional temporal NDVI profiles was performed assuming a linear regression model. Coefficient of determination R^2 and Root Mean Squared Error (RMSE) were computed to assess the models' errors and to analyze the effect of the aggregation on the different crop masks. Results indicated an improvement in yield estimation when using annual crop group-specific indicators with respect to generic and static products. Although not homogeneously distributed throughout Europe, advantages have been highlighted both in terms of accuracy and timeliness of the prediction. In most regions, the introduction of annual masks allowed to reduce RMSE values by 0.3 t/ha and advanced forecast times by up to 30 days. The added value in the use of annual crop group-specific masks concerned the two European most cultivated crops, namely soft wheat and grain maize.

1. Introduction

Increasing worldwide demand for agricultural products, combined with inter-annual fluctuations of global production are important drivers for grain price volatility on agricultural markets. Moreover, food security challenges are even more important to address in an era of global climate change (Ansarifar et al., 2021). In this context, yield forecasting plays a major role in anticipating production anomalies, allowing well-informed and timely policy actions, preventing food crises, avoiding market disruptions, and contributing to overall increased food security (López-Lozano et al., 2015a). Crop yield prediction is a fundamental research topic in agricultural and environmental sciences.

* Corresponding authors.
E-mail addresses: giulia.ronchetti@ext.ec.europa.eu (G. Ronchetti), giacinto.manfron@ec.europa.eu (G. Manfron), christof.weissteiner@ec.europa.eu (C.J. Weissteiner), lorenzo.seguini@ec.europa.eu (L. Seguini), luigi.nisini-scacchiafichi@ext.ec.europa.eu (L. Nisini Scacchiafichi), lorenzo.panarello@ext.ec.europa.eu (L. Panarello), bettina.baruth@ec.europa.eu (B. Baruth).

https://doi.org/10.1016/j.compag.2023.107633
Received 14 September 2022; Received in revised form 20 December 2022; Accepted 7 January 2023
Available online 19 January 2023

This focuses on understanding how harvestable plant biomass is determined by genotype (G), environment (E), management (M), and their interactions (G × E × M) (Boschetti et al., 2017; Hipólito de Sousa et al., 2018; Cooper et al., 2020). Timely and reliable crop yield forecasts play an important role in supporting national and international agricultural and food security policies, stabilizing markets and planning food security interventions in food-insecure countries (Becker-Reshef et al., 2020).

Governments and public authorities, as the European Union (EU), have a strong need and mandate to anticipate crop production losses (Meroni et al., 2021). For them crop yield forecasts and crop production estimates are necessary to provide decision makers with timely information for rapid and effective response during the growing season. This information is obtained thanks to Decision Support Systems (DSS) providing crop yield forecasts, as in the case of the European Commission's Joint Research Centre (EC-JRC) in-house MARS (Monitoring Agricultural Resources) Crop Yield Forecasting System (MCYFS, Van der Velde et al., 2019) or the yield forecasting program of the United States National Agricultural Statistics Service (NASS: NASS, 2006). Such systems are often based on regressive estimation models developed by means of official yields and agro-environmental variables, computed at the time of the forecast (Fritz et al., 2019). The relationship usually relies on historical series of statistical yields and on one or more regressors, selected among meteorological data, crop simulation model or satellite-derived indicators (Basso et al., 2013; López-Lozano et al., 2015b). The fitting between estimators and crop yields is highly variable across the agronomical seasons and the reliability with which these variables infer on yields depends, among many other factors, also on their aggregation in the space domain (Pianosi et al., 2016); for example, on the quality and representativeness of the utilized agricultural land cover masks (Zhang et al., 2019; Pérez-Hoyos et al., 2020).

Earth Observation (EO) data, exploited in form of vegetation or biophysical indicators, represent a key variable in the design of yield predictive models (Basso et al., 2013; Kouadio et al., 2014; Zhang et al., 2019). The added-value concerns the sensitivity they have with respect to the agro-ecosystems combination of G × E × M factors and agro-practices (Bandyopadhyay et al., 2014; Hatfield and Walthall, 2015). Nevertheless, EO indicators show limits in their applications due to land cover maps availability (pixel selection - Liu et al., 2019; Zhang et al., 2019) but also due to the bias introduced when mixed-pixels are considered (low-resolution bias - Boschetti et al., 2004). Improving the representativeness of satellite-based indicators is a challenge for operational yield forecasting systems, especially when they are operative at a wide geographical scale (Atzberger, 2013). Remote sensing (RS) indicators in this context are usually spatially aggregated using crop masks resulting from automatic (supervised or unsupervised) classification methodologies (Rojas et al., 2011). Crop (binary) masks are therefore exploited to focus on those pixels in the satellite images belonging to the targeted agricultural land cover and are of fundamental importance to improve spatial aggregation of EO indicators to increase model prediction accuracy (Baruth and Kucera, 2006; Mkhabela et al., 2011).

Ideally, a crop mask should be differentiated by specific crop type and updated within the season to address agricultural land use changes (Zhang et al., 2019). However, due to the lack of timely availability of ground truth data within the crop growing season, it is not normally practical to get up-to-date annual crop masks within the crop season (Waldner et al., 2015; Davidson et al., 2017). As a result, the most used crop masks in yield forecasting systems provide static and generic land cover information without considering annual variations or crop groups (Baruth and Kucera, 2006; Mkhabela et al., 2011; Chipanshi et al.,2015). A recent study proposed a semi-automatic approach to identify crop group-specific pure pixels (i.e., winter and spring crops and summer crops) at European scale, based on the implementation of a regional Gaussian Mixture Model (GMM) on MODIS–NDVI time series analysis (Weissteiner et al., 2019). Such input could be used to improve the predictability of crop monitoring and yield forecasting applications, as it introduces a new and more detailed information layer to agricultural land cover in Europe. In a DSS-type operative environment and with reference to the MCYFS, the implementation of this method could bring a twofold advantage: on the one hand it would allow to benefit from more specific agricultural land cover masks, in particular allowing the transition from generic layers (e.g. arable areas) to more specific ones (e.g. summer crops); on the other hand it would allow to implement the transition from general (static) selective masks to annual (dynamic) mask, avoiding therefore the implicit assumption of an invariant agricultural land use in the time domain.

The aim of this work is to test the contribution of crop group-specific and dynamic information for regional yield estimation in Europe. We propose compared correlation analyses between yield data and regional temporal NDVI profiles extracted from several land cover masks to test if crop group-specific RS time series and even annual crop group-specific RS time series fit more than generic RS time series, when correlated with crop yields. Results are discussed in view of their applicability to agricultural DSS monitoring systems at regional level, with particular attention to the improvement of crop yield predictability along the agronomical season. The analyses discussed in this contribution focused on three main goals: (i) test the use of annual crop group-specific NDVI time series to estimate crop yields; (ii) analyse results distribution at regional scale all over EU; and (iii) evaluate the potential benefit in terms of yield prediction accuracy and timeliness for two different crop groups: winter/spring and summer crops.

2. Materials and methods

2.1. Study area

The study area included all current EU Member States, except for Finland and Malta due to a lack of arable land areas. The area of interest covered a wide extent, ranging from 35° N to 60° N, and included different climatic conditions and agro-management practices. Therefore, to account at regional level for both climatic and technological characteristics, we performed the analysis at regional level, considering the administrative units from the Nomenclature of Territorial Units for Statistics at level 2 (NUTS-2 – Eurostat, 2018). Wherever possible, we selected for each EU Member State the five regions with the most prominent shares of arable land, according to the Corine Land Cover (CLC) 2018 agricultural classes (Kosztra et al., 2017). The final study area included 97 NUTS-2 regions, representative for 72% of the arable land in the EU and, at country level, representative for 85% (on average) of the national arable lands. The selected NUTS-2 regions are reported in Table 1 and shown on the map in Fig. 1.

2.2. Reference data

Regional (NUTS-2) crop yield statistics were collected from official databases (including Eurostat and the National Statistics Offices (Ronchetti et al., 2022)) for the 2003–2019 reference period and used as reference data for computing regressions. Considering the variety of crops cultivated at European level, yield statistics were retained only for the prevalent crops inside each region. In line with Weissteiner et al. (2019), the main agricultural crops were first divided into two crop groups, namely Winter and Spring Crops (WSpCs) and Summer Crops (SCs). Then, for each of the selected regions, the most representative crop of each crop group was chosen in accordance with the average extent of cultivated area. Among WSpCs, soft wheat prevailed (75 out of 97 regions), while grain maize and potato ranked first and second among SCs (48 and 28 out of 97 regions, respectively). The crops chosen for the analysis in each NUTS-2 region are reported in the maps of Fig. 2.

The length of each reference yield time series varied according to the availability of statistical data for that region and crop in the official database, but in most cases (65%), they covered the whole period of analysis (i.e., from 2003 to 2019). To spot significant trends in regional

Table 1
The NUTS-2 regions chosen for the analyses.

Country	Selected NUTS-2 regions
AT	AT11 (Burgenland), AT12 (Niederösterreich),AT21 (Kärnten), AT22 (Steiermark), AT31 (Oberösterreich)
BE	BE23 (Prov. Oost-Vlaanderen), BE25 (Prov. West-Vlaanderen),BE32 (Prov. Hainaut), BE33 (Prov. Liège), BE35 (Prov. Namur)
BG	BG31 (Severozapaden), BG32 (Severen tsentralen),BG33 (Severoiztochen), BG34 (Yugoiztochen), BG42 (Yuzhen tsentralen)
CY	CY (Kýpros)
CZ	CZ02 (Střední Cechy), CZ03 (Jihozápad),CZ05 (Severovýchod), CZ06 (Severovýchod), CZ07 (Střední Morava)
DE	DE40 (Brandenburg), DE80 (Mecklenburg-Vorpommern),DE94 (Weser-Ems), DEE0 (Sachsen-Anhalt), DEF0 (Schleswig-Holstein)
DK	DK01 (Hovedstaden), DK02 (Sjælland),DK03 (Syddanmark), DK04 (Midtjylland), DK05 (Nordjylland)
EE	EE (Eesti)
EL	EL51 (Anatoliki Makedonia, Thraki), EL52 (Kentriki Makedonia),EL53 (Dytiki Makedonia), EL61 (Thessalía), EL64 (Sterea Ellada)
ES	ES24 (Aragón), ES41 (Castilla y León),ES42 (Castilla-La Mancha), ES43 (Extremadura), ES61 (Andalucía)
FR	FRB0 (Centre - Val de Loire), FRE2 (Picardie),FRF2 (Champagne-Ardenne), FRG0 (Pays de la Loire), FRI3 (Poitou-Charentes)
HR	HR03 (Jadranska Hrvatska), HR04 (Kontinentalna Hrvatska)
HU	HU21 (Közép-Dunántúl), HU22 (Nyugat-Dunántúl),HU23 (Dél-Dunántúl), HU32 (Észak-Alföld), HU33 (Dél-Alföld)
IE	IE04 (Northern and Western), IE05 (Southern), IE06 (Eastern and Midland)
IT	ITC4 (Lombardia), ITF4 (Puglia),ITG1 (Sicilia), ITH3 (Veneto), ITH5 (Emilia-Romagna)
LT	LT01 (Sostinės regionas), LT02 (Vidurio ir vakarų Lietuvos regionas)
LU	LU (Luxembourg)
LV	LV (Latvija)
NL	NL11 (Groningen), NL13 (Drenthe),NL23 (Flevoland), NL33 (Zuid-Holland), NL34 (Zeeland)
PL	PL41 (Wielkopolskie), PL61 (Kujawsko-Pomorskie),PL62 (Warminsko-Mazurskie), PL81 (Lubelskie), PL92 (Mazowiecki regionalny)
PT	PT11 (Norte), PT15 (Algarve), PT16 (Centro),PT17 (Área Metropolitana de Lisboa), PT18 (Alentejo)
RO	RO21 (Nord-Est), RO22 (Sud-Est),RO31 (Sud - Muntenia), RO41 (Sud-Vest Oltenia), RO42 (Vest)
SE	SE12 (Östra Mellansverige), SE21 (Småland med öarna),SE22 (Sydsverige), SE23 (Västsverige), SE31 (Norra Mellansverige)
SI	SI03 (Vzhodna Slovenija), SI04 (Zahodna Slovenija)
SK	SK01 (Bratislavský kraj), SK02 (Západné Slovensko),SK03 (Stredné Slovensko), SK04 (Východné Slovensko)

yield statistics (Wu et al., 2007), we performed a Mann-Kendall test with a significance level equal to 1% (Mann, 1945; Kendall, 1975). Where required, crop yield statistics were detrended assuming a linear trend model. A significant increasing trend was identified and removed for 17 regions in the WSpCs group and for 35 regions in the SCs group. Regions with an identified trend in yield statistics are highlighted in Fig. 2 as hatched areas.

2.3. Regional temporal NDVI profiles

Starting from MODIS daily products at 250 m spatial resolution, regional temporal NDVI profiles were generated by aggregating pixels using different crop masks. For each region and year, five different NDVI average profiles were extracted. A detailed description of the main processing steps and of the applied masks is following.

2.3.1. MODIS NDVI time series processing

A collection of reference temporal NDVI profiles was derived for every selected region and year using MODIS daily product at 250 m spatial resolution. More specifically, the MOD09GQ product (250 m resolution, Terra platform) from the collection 6 was adopted in this study, along with MOD09GA (500–1000 m resolution) for retrieving additional information on quality rating and satellite acquisition geometries (Vermote et al., 2015; Weissteiner et al., 2019). The datasets were provided by NASA LP DAAC at the USGS EROS Center and represented daily, global, and atmospherically corrected surface reflectance values in the red and near-infrared (NIR) domains. Data processing routines were designed and performed both in cloud computing, via the Google Earth Engine application (Gorelick et al., 2017), and at desktop level via Python (Van Rossum and Drake, 2009). Following the approach of Weissteiner et al. (2019), MODIS data were first limited to cover the extent of the study area and acquisition period corresponding to the seasonal growing cycle of the main agricultural crops in Europe, from February to October. Then, noisy data were filtered out according to the quality flags provided in the MODIS product quality flags and NDVI time series at 250 m spatial resolution were computed from bands 1 and 2 of the MOD09GQ.006 product, as indicated in Rouse et al. (1974):

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \quad (1)$$

Finally, a fifth degree polynomial was applied to each pixel time series for smoothing and producing yearly NDVI time series. More details on the filtering and quality flags used in this study can be found in Weissteiner et al. (2019).

2.3.2. Retrieving regional temporal NDVI profiles

To assess the effects of crop masks on the accuracy of yield estimation, regional annual NDVI profiles were retrieved by applying different crop masks. Five masks were considered in this analysis: a generic arable land mask and two crop group-specific masks for both WSpCs and SCs groups. The arable land mask was computed from the agricultural classes of the CLC 2018, while the crop group-specific masks were generated from the datasets of crop group-specific pure pixels presented in Weissteiner et al. (2019).

The adopted masks are illustrated hereafter:

- ArLand: generic arable land mask, computed from the CLC 2018. It included not-irrigated (class 2.1.1) and permanently irrigated (class 2.1.2) arable land areas (Kosztra et al., 2017). This mask provided information that was characterizing the whole study area, but in a static and generic way. Specifically, it did neither consider inter-annual variations of the arable area extension nor distinguished crop groups.
- HistWSpCs: historical crop group-specific mask for the WSpCs group. It represented a (pixel-based) probability of more than 50% to identify a pure pixel of WSpCs in the 2003–2019 time period. This mask was computed from the pure pixels of WSpCs detected each year (Ronchetti et al, 2021). It provided static crop group-specific information, without considering annual variability.
- HistSCs: historical crop group-specific mask for the SCs group. It was the equivalent of the HistWSpCs mask, for the SCs group.
- YearWSpCs: annual crop group-specific mask for WSpCs group. It represented the pure pixels for the WSpCs group, detected in a specific year. This mask changed annually, depending on the location of the pure pixels identified each year. It provided dynamic crop group-specific information that accounted for both annual variability and specific characteristics of crop groups.
- YearSCs: annual crop group-specific mask for SCs group. It was the equivalent of the YearWSpCs mask, for the SCs group.

To retrieve average regional temporal NDVI profiles, a dedicated Python (version 3.7; Van Rossum and Drake, 2009) routine was developed. This script used as inputs the MODIS NDVI time series, the different crop masks, and the European regional borders. It provided as output the desired average NDVI profiles for each region of interest (n = 97), year (n = 17) and mask (n = 5). The procedure involved a selection of pixels from MODIS imagery based on two criteria, namely pixels belonging to a specific crop mask and included within the borders of the target region. The selected pixels were then aggregated to compute regional statistics, including mean and standard deviation values, thus

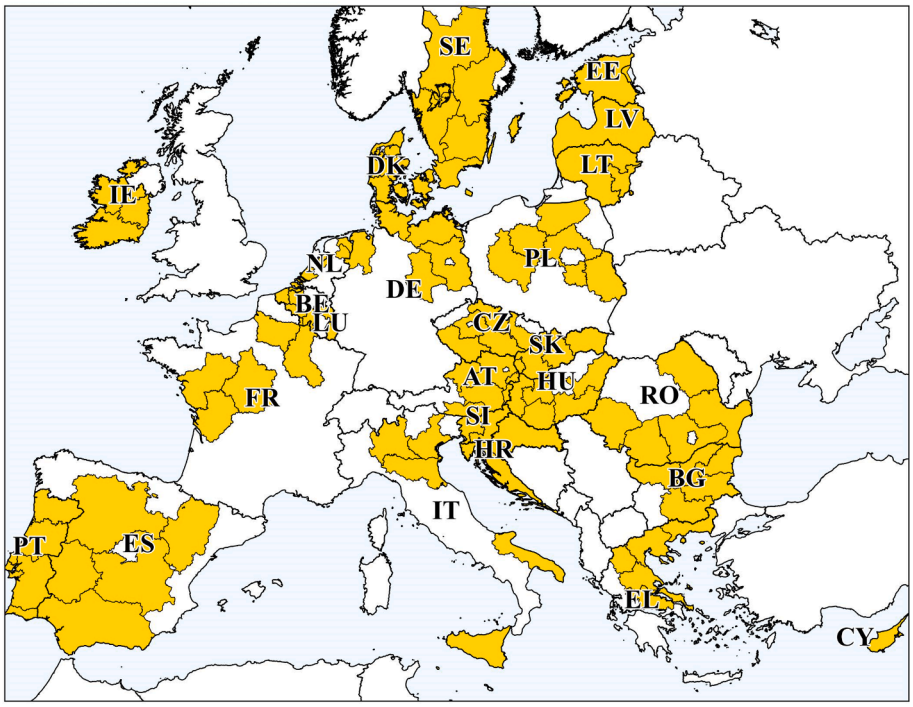


Fig. 1. The European NUTS-2 regions (in yellow) selected as study area.

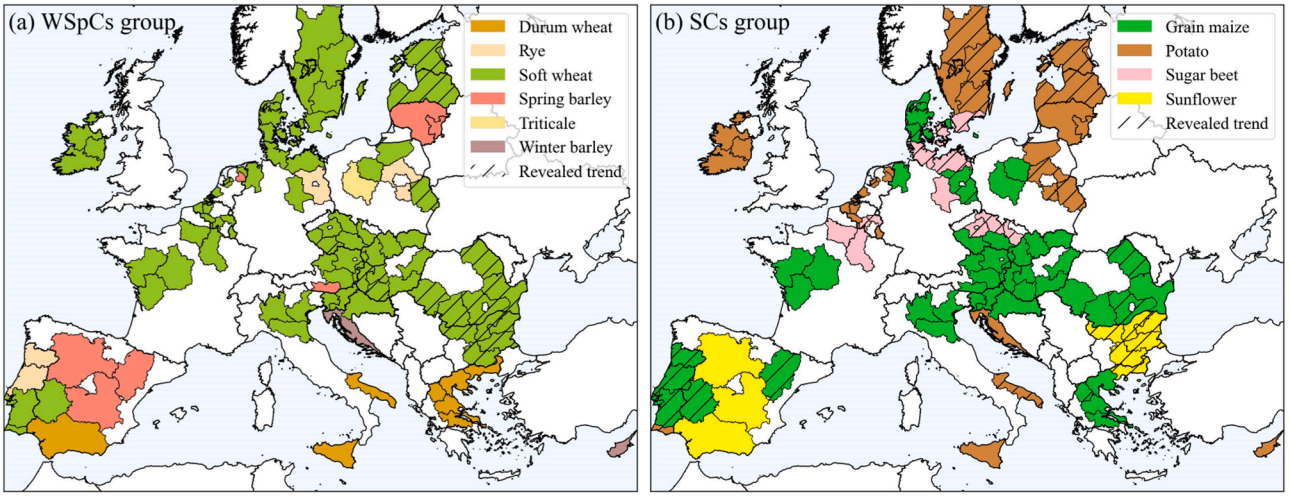


Fig. 2. The prevalent crops selected inside each NUTS-2 region for (a) WSpCs and (b) SCs. Hatched areas display a significant trend in the crop yield statistics.

providing the desired regional NDVI profiles.

2.4. Regression model

A correlation analysis was performed between reference yield data and regional temporal NDVI profiles assuming a linear regression model. For each region and mask, annual yield values were compared with the average NDVI values of the corresponding region and year. To assess the evolution in time and find the best time of the year to compute regressions, we did not consider only NDVI peak values but all NDVI values, starting from the day of the year (DOY) 60 (i.e., beginning of March) up to DOY 270 (i.e., end of September). However, to reduce the number of models, NDVI values were sampled every ten days, totaling 22 models for each region and mask.

NDVI profiles aggregated with HistWSpCs mask and YearWSpCs mask were tested only with respect to yield values for the WSpCs group and accordingly, NDVI profiles aggregated with HistSCs mask and YearSCs mask with respect to yields for the SCs group. As far as NDVI profiles aggregated with ArLand mask concerned, correlations were computed for both the crop groups.

The coefficients of determination R^2 (Eq.2) were calculated to assess the performance of each relationship, together with the respective p-value to estimate its significance. The resulting metrics obtained from the correlations between yield and NDVI aggregated with the different crop masks were compared and analyzed.

$$R^2 = \sum_{t=1}^{n}(F_t - \overline{O})^2 / \sum_{t=1}^{n}(O_t - \overline{O})^2 \quad (2)$$

Where F_t are yield values predicted from the model for each specific year t, O_t are the observed yield values from the reference dataset and $\overline{O}$ are their average values. For both crop groups and for all the five crop masks used in the aggregation phase, we evaluated only relationships that returned a significant and high correlation, having set the level of

significance (α) to 5% and a threshold on R^2 values equal to 0.4. This threshold was arbitrarily chosen as representative of a strong correlation. To assess both the accuracy and the timeliness of the prediction, for each region and crop group we selected the aggregating mask providing the maximum R^2 value (above the threshold) and compared R^2 variability along the season.

In addition, since this work is included in the context of a crop yield forecasting system, the correlation results obtained with NDVI profiles aggregated with the different masks were analyzed and compared also in terms of the forecasting errors, namely the Root Mean Square Error RMSE (Eq.3):

$$RMSE = \sqrt{\sum_{t=1}^{n}(O_t - F_t)^2/n} \quad (3)$$

To compare forecasting errors, the differences of the lowest RMSE, the first-in-time significant RMSE and their temporal occurrences were considered.

3. Results

This section introduces the research results in form of graphs and maps. To improve in readability, we chose a unique colour for displaying data resulting from each of the (five) compared crop masks. In particular: grey (ArLand), red (HistWSpCs), blue (HistSCs), orange (YearWSpCs) and cyan (YearSCs). In the first subsection (§3.1), NDVI time series resulting from the different spatial aggregation crop masks are shown. The second subsection (§3.2) illustrates correlation results considering the whole study area, while a focus on regional comparisons is provided in §3.3 and §3.4, depicting the best retrieved R^2 value and R^2 variability along the season, respectively. Finally, §3.5 is dedicated to the analyses on yield estimation errors.

3.1. NDVI spatial aggregation

The regional temporal NDVI profiles vary as function of the mask applied during the spatial aggregation phase. Fig. 3 shows eight examples of NDVI profiles, extracted from selected regions and years among the study area. Solid lines represent NDVI mean values, while error shadows refer to +/- one standard deviation. Given a generic region of interest, NDVI profiles differ according to (i) the applied crop mask; (ii) the number of pixels considered in the aggregation phase; (iii) the temporal occurrence of the NDVI peak; (iv) the NDVI value at the seasonal peak; and (v) the variability of NDVI values (i.e., standard deviations). Examples are taken from different EU regions to show the consistency of masks aggregation effect across the study area and across years. NDVI profiles suggest homogeneous distribution of crop groups in the arable land area when a finer delineation is observed in the transition from ArLand to crop group-specific regional profiles (e.g., Fig. 3a and Fig. 3b). The predominance of one specific crop group in a region can be inferred when the ArLand and crop group regional profiles tend to overlap (e.g., Fig. 3c, Fig. 3d, Fig. 3e, Fig. 3f, Fig. 3g and Fig. 3h). The examples in Fig. 3g and Fig. 3h, are representative for the crop group

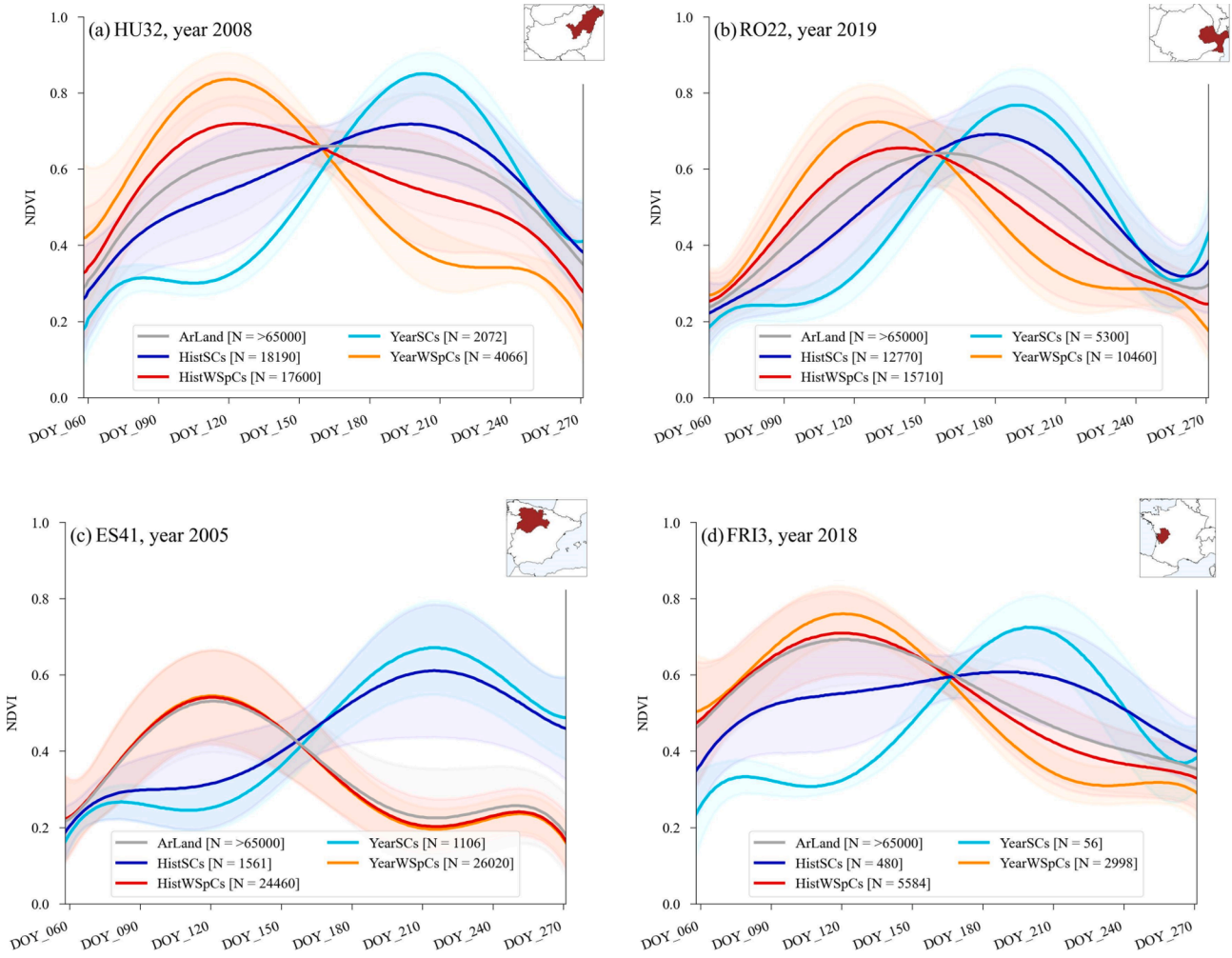


Fig. 3. Temporal NDVI profiles extracted at regional level using five different crop masks. Solid lines display average NDVI values, shadows highlight +/- one standard deviation. The number of aggregated pixels is reported in brackets into the legend. Orientation maps are provided for georeferencing each example.

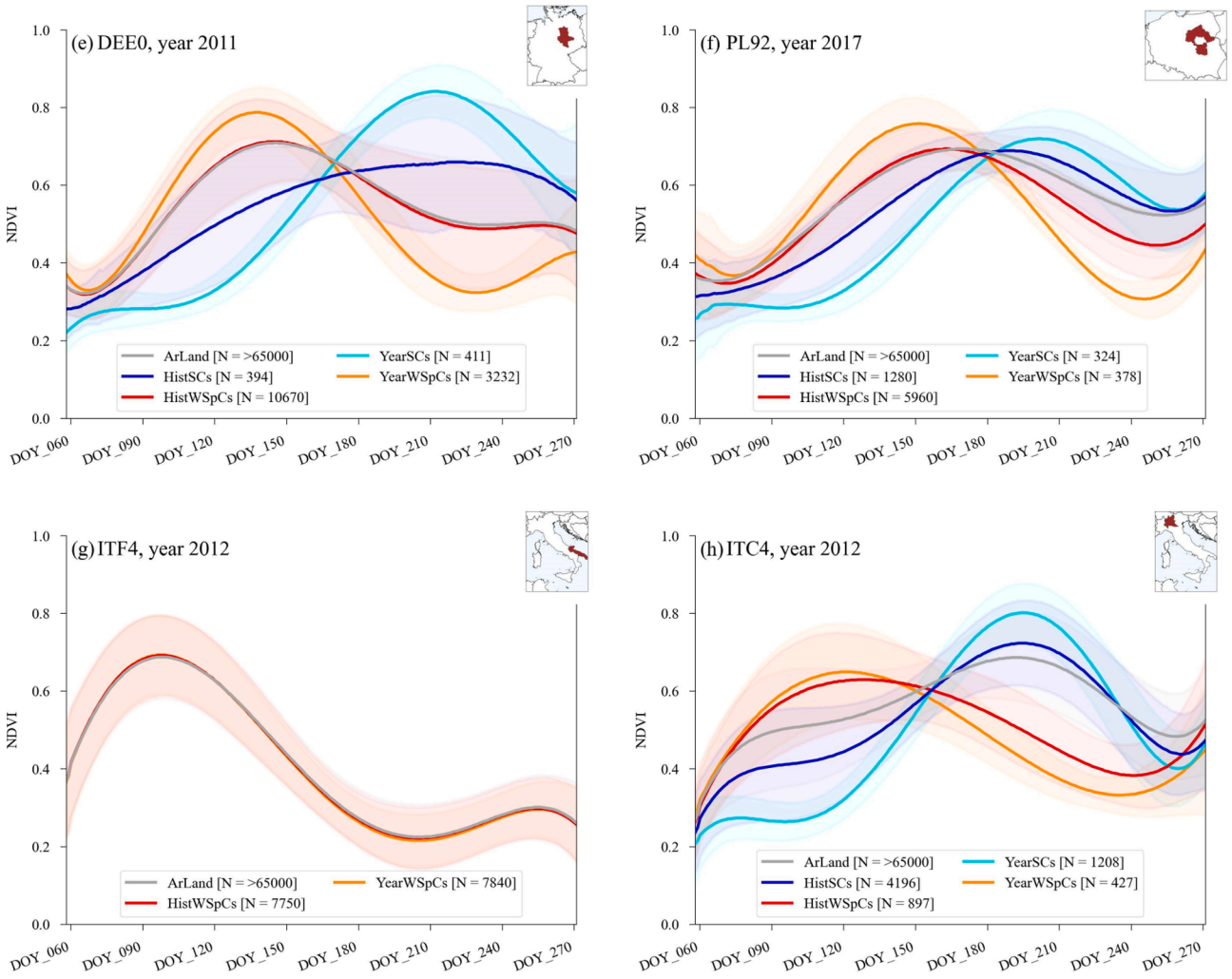


Fig. 3. (*continued*).

variability among regions of the same country for the year 2012. A high prevalence of WSpCs can be depicted in the NDVI temporal patterns in Fig. 3g while SCs prevail in the temporal patterns of Fig. 3h.

3.2. Results at continental scale

Considering the (97) regions accounted in this study, the (5) compared crop masks, the (2) investigated crop groups and the selected time stamps (22, from DOY 60 to DOY 270); a total of 12,804 regression models were computed. Among these, 2906 (22.7%) showed significant correlation results (p-value $< \alpha$), for a level of significance α set to 5%. The distribution and the number of significant correlations spread region-wise as indicated in the map of Fig. 4. Although not evenly distributed across the study area, significant relationships were observed for almost all the regions under analysis (94 out of 97). On average, 30 significant models resulted for each region. The number of significant models mostly depend on the extension of arable land inside each region (e.g., regions with higher number of significant correlations were found in Spain) and the considered prevalent crop (e.g., the lowest number of significant correlations were found for durum wheat and potato). In a large part of the accounted regions, i.e. in 38 out of 97, significant correlations are in the range 25–50.

Although the numbers of significant and high correlations were quite homogeneous considering the different aggregation masks, their distributions over time differed. The histograms mapping the distribution of the significant correlations and high correlations over time are provided

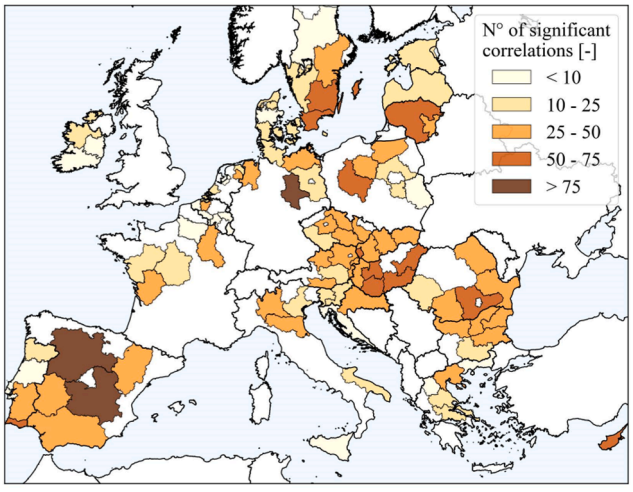


Fig. 4. Distribution of significant correlations (p-value < 0.05) by region.

in Fig. 5. It can be noticed that significant relationships were mainly localized in a time span ranging around the NDVI peaks, confirming what was already highlighted in other contributions (Skakun et al., 2017; Nagy et al., 2018; Liu et al., 2020; Meroni et al., 2021; Shammi and Meng, 2021). As a matter of fact, the largest number of significant

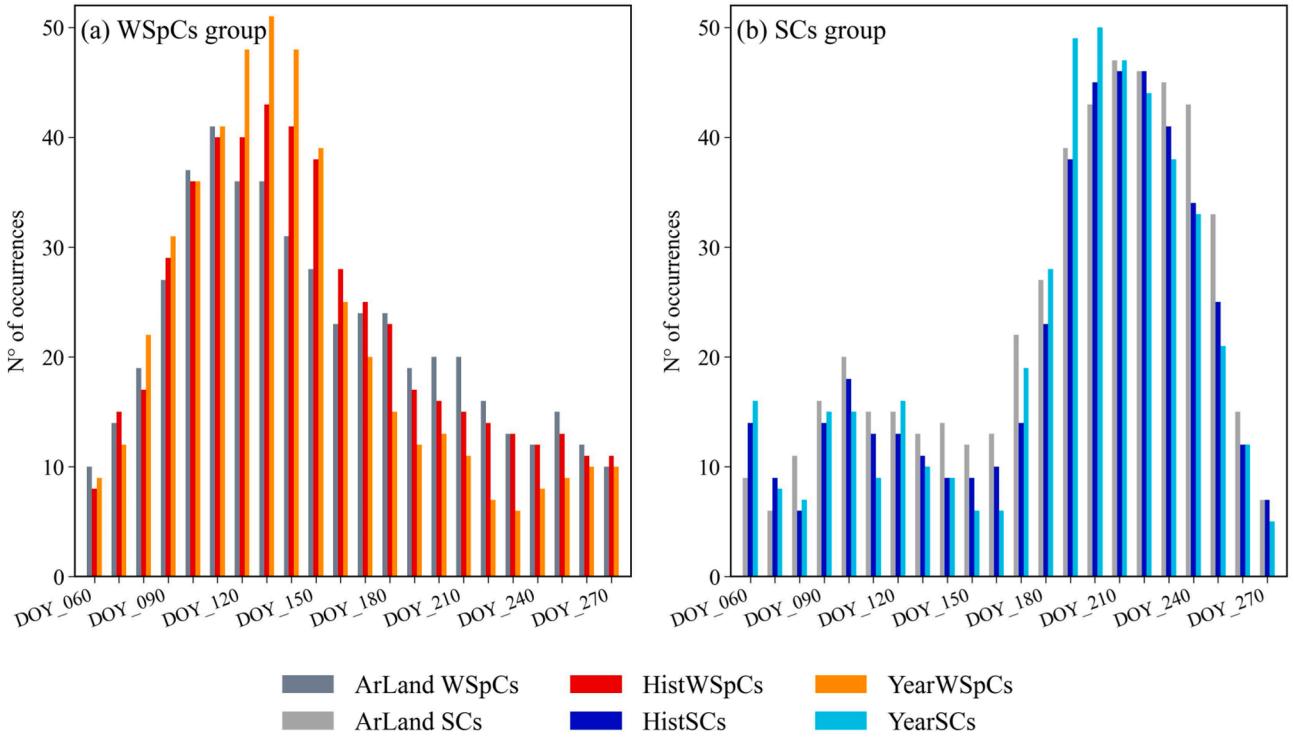


Fig. 5. Distribution of significant correlations over time: a) WSpCs group, b) SCs group.

correlations for the WSpCs group was found approximately from DOY 90 to DOY 150 (Fig. 5a), while for SCs group between DOY 190 and DOY 240 (Fig. 5b). This was observed regardless of the mask considered.

Regarding the aggregating masks, for WSpCs group, the number of significant correlations computed with NDVI values aggregated using the annual mask is higher than the other in the range from DOY 90 to DOY 150 and lower out of this time span (Fig. 5a). The same for the SCs group, where the number of significant correlations computed with NDVI values aggregated using the annual mask resulted higher than the others between DOY 180 and DOY 210 and lower out of this interval (Fig. 5b).

Fig. 6 shows pairwise comparisons of correlations between NDVI values and crop yield, when different crop masks are used to aggregate NDVI data at regional level. Histograms were drawn to represent the total amount of significant correlations (i.e., total number of comparisons – black bars) and the times an aggregation mask returned higher R^2 values than the opposing mask (colored bars). The same results are also expressed in terms of relative values by means of continuous and dashed lines, referring to the right y-axes of each pane. The histograms confirmed once more a pronounced number of significant correlations around NDVI peaks (proxy of crop headings), more markedly for SCs group than for WSpCs. By comparing correlation results, the R^2 values from YearWSpCs aggregated models resulted higher than the outcomes obtained by the other models. Similar results were observed for the SCs group, where YearSCs aggregated indicators returned greater R^2 values with respect to both the ArLand and HistSCs aggregated data. While R^2 values with the use of ArLand tended to prevail over HistSCs.

The distribution of R^2 average values over time for the whole dataset is shown in Fig. 7. Only significant correlations were included in the results. Overall, the R^2 average values were in the range 0.3–0.65. An increase in R^2 average values was again highlighted for DOYs close to the agronomical heading of the accounted crop group, except for some high R^2 values resulting from the correlation between yield and NDVI aggregated using ArLand and HistWSpCs mask, for WSpCs group. R^2 average values were higher in the SCs group than in the WSpCs group, when compared with reference to their respective agronomical heading periods. NDVI profiles aggregated using YearWSpCs mask resulted in R^2 average values equal or slightly higher than the other masks in the period from DOY 90 to DOY 130 and lower for the remaining DOYs (Fig. 7a); similarly, the use of YearSCs mask allowed to obtain higher R^2 average values from DOY 180 to DOY 220 (up to greater than 0.15) and lower outside this interval (Fig. 7b). However, in this analysis R^2 results were mitigated by the fact that we were considering different regions and several crops together. In the next sections (§3.3, §3.4), the comparisons will instead focus on individual regions and/or crops.

3.3. Results at regional scale

A more detailed analysis focused instead only on the highest correlation results, where models returning R^2 values below the threshold value (i.e., 0.4) were removed. The considered models decreased to 1410 (11%), and their repartitions according to the original regressors are illustrated in Table 2. Comparisons between NDVI spatial aggregation masks were made region by region, identifying the mask returning the maximum R^2 value along the reference agricultural time window. Results are shown in Fig. 8. For WSpCs group (a), YearWSpCs prevailed in 33.0% of the regions, 32 out of 97, while ArLand and HistSCs ranked second and third (15.5% and 13.4%, respectively). In 38.1% of the accounted regions, correlations were not significant or with R^2 values below the imposed threshold. The regions with a prevalence for YearWSpCs mask were fairly distributed in the study area, particularly in central and eastern Europe, including Czechia, Slovakia, Hungary, Romania and Bulgaria, in addition to some regions in Spain, France, Germany and Poland. Considering instead the SCs group, YearSCs prevailed in 42.3% of the regions, 41 out of 97, followed by ArLand and HistSCs (16.5% and 10.3%, respectively). In 30.9% of the regions were no significant or high correlations observed. A predominance of YearSCs mask was observed across the whole EU, specifically in Central and Eastern Europe, including some important agricultural areas of Europe, such as central-western France, northern Italy and eastern Romania.

Results previously illustrated in the maps of Fig. 8 accounted for all the regions and all the crops of a crop group. To better evaluate the

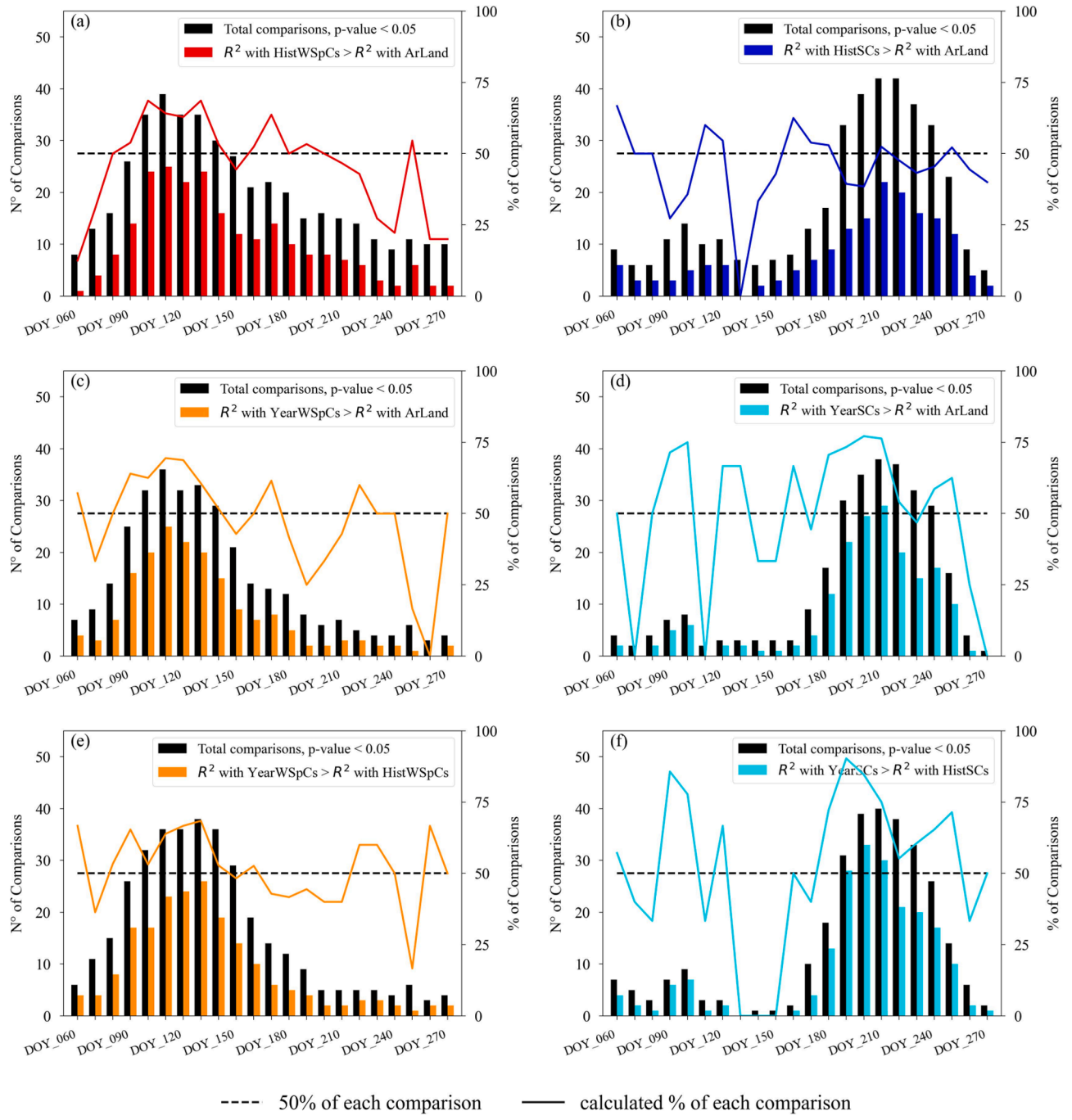


Fig. 6. Pairwise comparisons of significant correlations between NDVI and crop yield, varying the aggregation masks. Total number of comparisons (y axis - left) and referring percentage (y axis - right). a) HistWSpCs vs ArLand, b) HistSCs vs ArLand, c) YearWSpCs vs ArLand, d) YearSCs vs ArLand, e) YearWSpCs vs HistWSpCs, f) YearSCs vs HistSCs.

effects of NDVI spatial aggregation masks for specific and representative crops, we repeated our analysis crop-wise, by selecting only the regions featuring a specific crop prevalence. Here, we report results for the most prominent crops at EU level and within the two crop groups: soft wheat and grain maize (Fig. 9). For soft wheat, out of 75 regions, we obtained in 37.3% the best results in terms of maximum R^2 with NDVI aggregated using YearWSpCs mask. Together with the contribution of HistWSpCs (14.7%), more than 50% of the regions were reached; thus highlighting the importance of crop group-specific information in correlating with yields for soft wheat in Europe. For grain maize, 48 regions were selected and in 43.8% of these the maximum R^2 value was observed with the aggregation by means of YearSCs mask.

3.4. Regional variability of R^2 values along the season

Correlation results in terms of R^2 values varied at regional level as function of time and of the spatial aggregation mask used for generating NDVI profiles. In Fig. 10, the distribution of R^2 values by mask in the time domain is shown for selected regions and prevalent cultivated crops. The reported examples spread over the study area and are representative for main EU agricultural regions. Overall, significant correlations (solid lines) were observed in a temporal window from DOY 90 to DOY 250. Further investigations are needed to address the significant correlations observed after DOY 220 for WSpCs group and before DOY 120 for SCs group, occurring off the reference crop growing

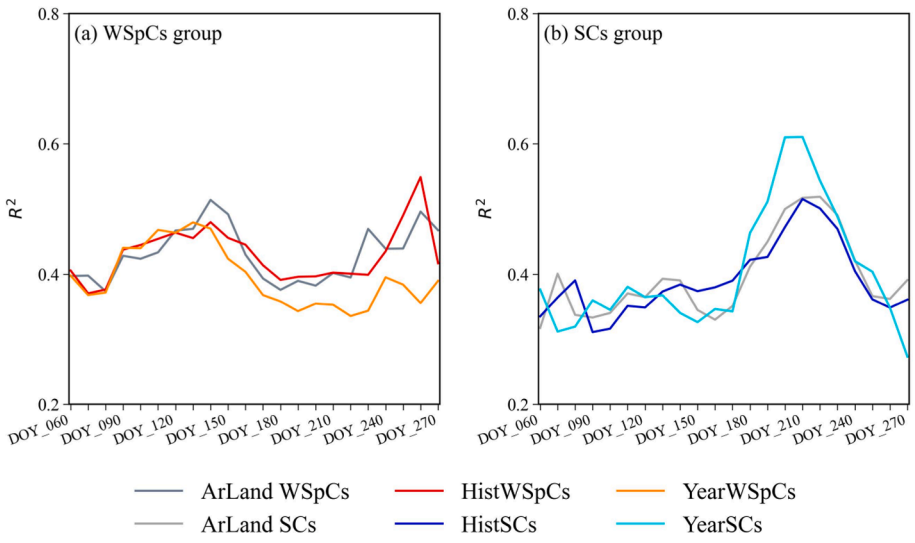


Fig. 7. Distribution of average R^2 values over time: a) WSpCs group, b) SCs group.

Table 2
Number of significant correlations (p-value < 0.05) and high correlations (R^2 > 0.4) by NDVI spatial aggregation mask.

Crop mask	N° significant correlations (p-value < 0.05)		N° high correlations (R^2 > 0.4)	
	WSpCs group	SCs group	WSpCs group	SCs group
Arland	487	511	244	235
HistWSpCs	505	–	237	–
HistSCs	–	457	–	204
YearWSpCs	483	–	225	–
YearSCs	–	463	–	265

window (i.e., after WSpCs harvesting and before SCs sowing periods). These correlations could be linked to the soil moisture condition in the pre-cropping season or to factors facilitating pests or weed spread during the incoming cropping season.

R^2 values for SCs group were generally higher than those obtained for the WSpCs group, with the exceptions of regions represented in Fig. 10d and Fig. 10f. This is most likely related to the high share of WSpCs cultivated area, much higher than SCs. In the other examples, the shares in the crop groups are almost equal, with only a slight prevalence of SCs in Fig. 10a and Fig. 10b, and of WSpCs in Fig. 10c and Fig. 10e. Comparing the outcomes of the correlations obtained using the different spatial aggregation masks, in most cases the R^2 values obtained with the annual masks were clearly over performing the others, both in terms of R^2 value and of early-in-time occurrence of high regression values. Only in Fig. 10d, we did not observe any improvements with the introduction of annual and crop group-specific masks, since in this region there is a large share of WSpCs group, as already pointed out in section 3.1 (Fig. 3c). In Fig. 10a, Fig. 10b, and Fig. 10c, the R^2 numerical improvements brought with the exploitation of annual crop group-specific masks were evident in both the crop groups, both in terms of R^2 values and earliness in the season. In Fig. 10a, R^2 values from arable land mask were close to 0.4, while these values raised up to 0.6 when the annual crop group-specific masks were introduced. In Fig. 10b, R^2 values reached 0.8 with the YearSCs mask. Furthermore, in these two regions, significant correlations for WSpCs group occurred up to 30 days in advance when annual crop group-specific masks were exploited. In Fig. 10c, there was a 20 days advance in significant correlations for SCs group when applying the YearSCs mask. In Fig. 10e and Fig. 10f, the improvements brought by the annual crop group-specific masks were evident only for one crop group. In Fig. 10e, the use of the YearWSpCs mask raised R^2 values for WSpCs group up to 0.5 around DOY 150, whereas in Fig. 10f for SCs group, we observed significant correlations and R^2 values greater than 0.4 around DOY 200 when the YearSCs mask was applied and ArLand and HistSCs masks did not return any significant correlations at any DOY. In all the selected examples, the general behavior of correlation results over time for the arable land mask was very similar to those of the two crop group-specific static masks.

3.5. Yield estimation errors: RMSE

To further assess the effects of the use of the different masks, we introduced the comparisons between the correlation results in terms of RMSE. In this section, only comparisons between annual crop group-specific masks and arable land mask were considered. The historical crop group-specific masks were left out of this analysis, as the results previously reported highlighted their worst performances.

For each region, we selected the minimum RMSE obtained from the correlations using annual crop group-specific and arable land aggregated indicators; we then assessed their differences both in terms of (i) error values and (ii) time-occurrence. Results are displayed in Fig. 11. Regarding RMSE residuals in t/ha (Fig. 11a and Fig. 11b), negative values (green colours) indicate regions where RMSE resulting from correlations with NDVI values aggregated on the annual masks are lower than those obtained with aggregation on the arable land. Positive values (red colours) reflect instead lower errors when applying the arable land mask. Yellow colours display regions where residuals computation was not applicable, because the use of at least one mask returned no significant correlations (i.e., p-value greater than 0.05 for all regression models). Overall, the number of regions reporting negative differences prevailed, thus highlighting an improvement in terms of forecasting accuracy when correlations were based on NDVI values aggregated using annual crop-specific masks. Particularly, for WSpCs group (Fig. 11a), we observed lower RMSE values with the aggregation on YearWSpCs in 34 regions, whereas in 17 regions RMSE values were higher. For SCs group (Fig. 11b), the use of YearSCs mask returned lower RMSE in 36 regions, and higher values in 13 regions. Larger differences between the use of the annual masks and the arable land mask were found for the SCs group compared to the WSpCs. In the latter, most of the residuals were in the range 0 to 0.1 t/ha (in absolute value) and there were no residuals in absolute value greater than 0.2 t/ha, while in the former there were residuals even in the class greater than 0.3 t/ha; thus, indicating non-neglectable changes in yield forecasting. Residuals computation was not applicable in 46 and 48 regions out of 97, for WSpCs and SCs group respectively. In particular, for SCs group, we found 20 regions where the residuals computation was not applicable, as

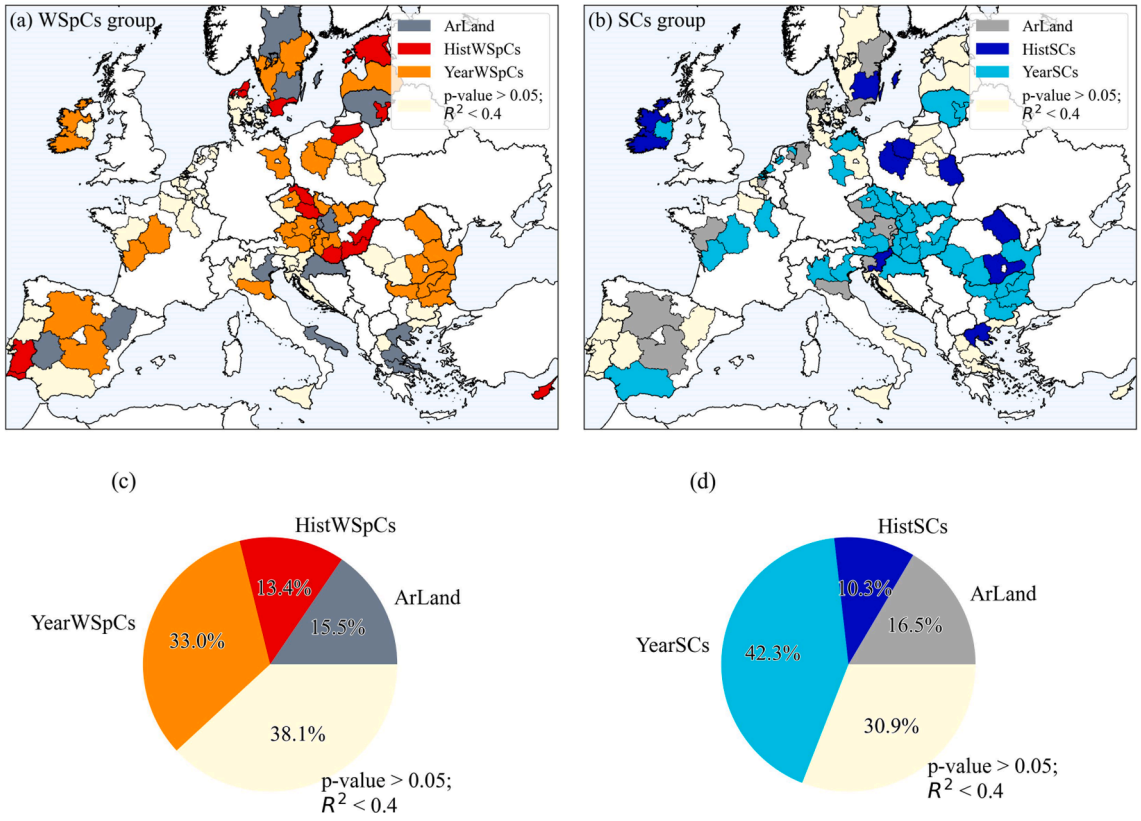


Fig. 8. High correlation results (R^2 greater than 0.4): (a,b) aggregation masks reaching the maximum R^2 within the agricultural window; (c,d) pie plots displaying the shares reached by each mask (total n = 97).

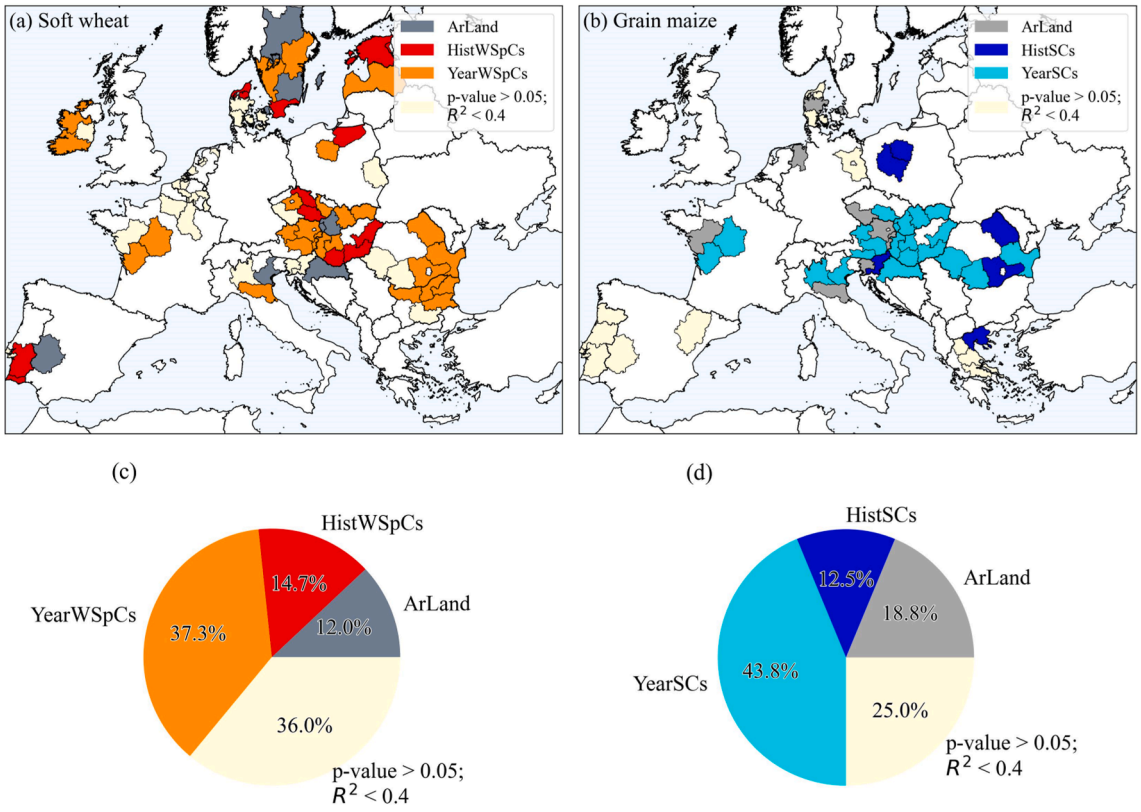


Fig. 9. High correlation results (R^2 greater than 0.4): (a, b) aggregation masks reaching the maximum R^2 within the agricultural window; (c, d) pie plots displaying the shares reached by each mask (total n = 75, WSpCs; total n = 48, SCs). Left (a, c): prevalent crop is soft wheat; right (b, d): prevalent crop is grain maize.

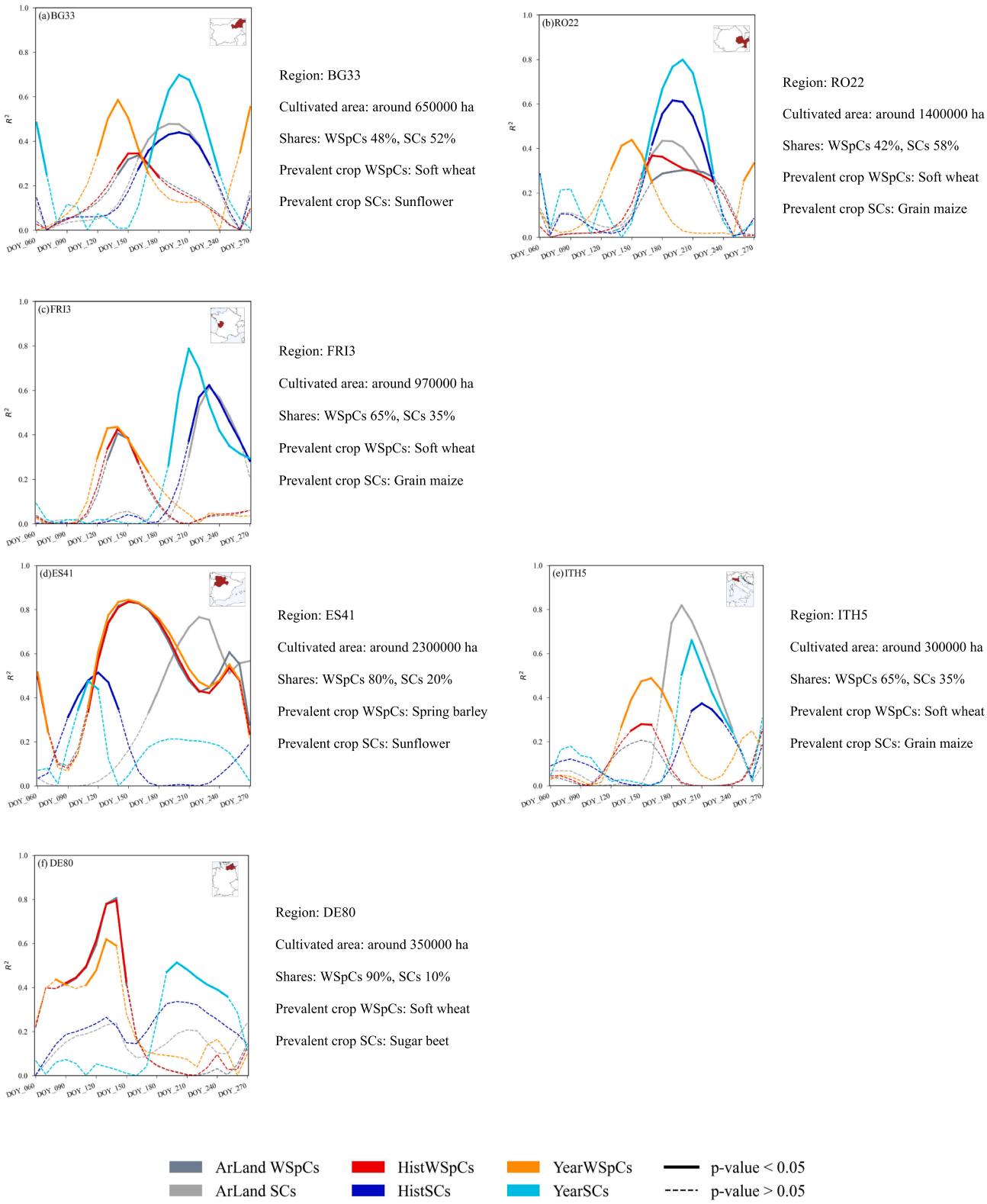


Fig. 10. Variability of R^2 values with respect to the time and the applied spatial aggregation mask. Example regions: a) BG33, b) RO22, c) FRI3, d) ES41, e) ITH5, f) DE80.

the outcomes of the use of YearSCs mask were missing. In these regions, the correlations between yield values and NDVI aggregated on YearSCs mask resulted non-significant or null, because of the lack of identified pure pixels. Specifically, it was not possible to compute any valid regression model in the northernmost countries, such as Estonia, Sweden, Denmark and Ireland, but also in the southernmost regions, including Greece and southern Italy, where summer crops cultivation is almost absent (Eurostat, 2018).

Regarding RMSE time-occurrence (Fig. 11c and Fig. 11d), green bars represent regions where correlations with NDVI values aggregated using

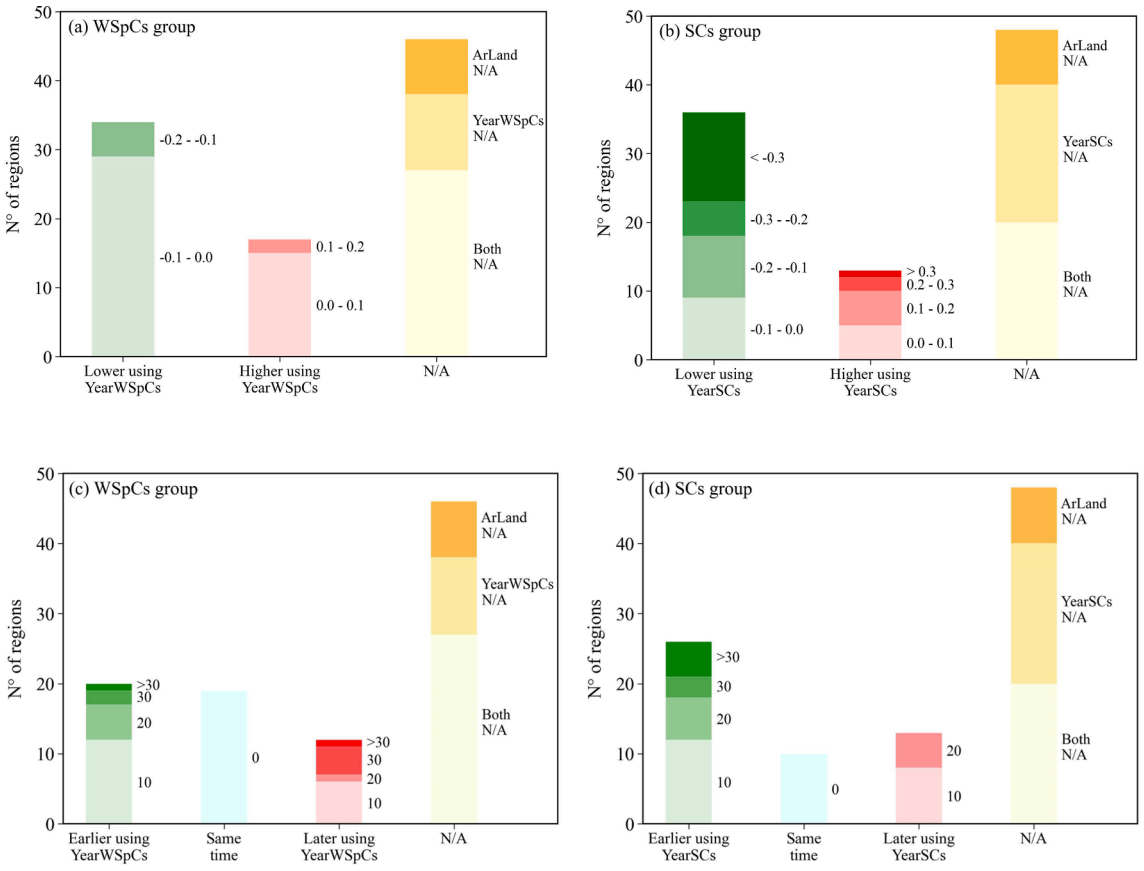


Fig. 11. Pairwise comparisons of the minimum RMSE: (a, c) YearWSpCs vs ArLand; (b, d) YearSCs vs ArLand. (a, b) RMSE residuals [t/ha]; (c, d) time-occurrence residuals [days].

annual masks returned the minimum RMSE value earlier than arable land aggregated indicators. Conversely, later occurrences are displayed with red bars. No changes in time-occurrence are shown with light-blue bars and yellow colours indicate regions where the comparison was not applicable (i.e., the same regions as in Fig. 11a and Fig. 11b). Overall, the number of regions where the minimum RMSE values were observed earlier-in-time prevailed, thus highlighting an improvement in terms of forecasting timeliness when correlations were based on NDVI values aggregated using annual crop-specific masks. For WSpCs group (Fig. 11c), the minimum RMSE with NDVI aggregated on YearWSpCs was returned earlier in 20 regions, at the same time in 19 regions, whereas later in 12 regions. In most cases, differences in time-occurrence were in the order of 10 days, but for few cases differences of 20 days and even more were observed. For SCs group (Fig. 11d), the use of YearSCs mask provided the minimum RMSE earlier in time in 26 regions, at the same time in 10 regions, and later in 13 regions. Residuals varied in time from 10 days to more than 30 days, but no cases of occurrences later than 20 days were registered.

In the context of a crop yield forecasting system, the capability to provide good predictions early in the season is of paramount importance (Basso and Liu, 2019; Shahhosseini et al., 2020; Ziliani et al., 2022). Therefore, we compared for each region also the values and time-occurrence of the first-in-time significant RMSE resulting from NDVI-yield correlations when applying annual crop group-specific masks and arable land mask (Fig. 12). Here, for WSpCs group (Fig. 12a), there was a slight prevalence of positive residuals (28 regions vs 23 regions). This indicates that overall correlations based on NDVI aggregated on YearWSpCs provided greater first-in-time significant RMSE values when compared to correlations based on NDVI aggregated on ArLand. However, RMSE variations were minimal and never exceeded 0.1 t/ha. For SCs group (Fig. 12b), negative residuals prevailed (green colours): in 30 regions the first-in-time significant RMSE was lower when applying YearSCs mask than ArLand, and greater in 19 regions. Regarding time-occurrence (Fig. 12c and Fig. 12d), we noticed again the added value brought by annual crop group-specific indicators, which returned significant RMSE earlier in the season or at most at the same time than arable land indicators. Particularly, for WSpCs group (Fig. 12c), in 27 regions significant RMSE were observed from 10 up to more than 30 days in advance, while only in 3 regions significant RMSE occurred later in the season. For SCs group (Fig. 12d), in 23 regions significant RMSE were found earlier when using YearSCs mask, and among these, in 12 regions the earliness is greater than 20 days.

4. Discussion

This section proposes a critical analysis on the obtained results, with close attention to already available literature contributions. The potential (qualitative and quantitative) added-value of annual crop group-specific masks for crop yield forecasting is discussed, with emphasis on the applicability for operational agricultural monitoring systems at regional level.

4.1. NDVI time series variability

Results provided in §3.1 support the reliability of EO time series for operational crop yield forecasting at regional level and show how yield predictive models can be improved by enhancing the specificity of NDVI time series, exploited in form of regressive inputs. These findings confirmed at European level the evidence brought by Zhang et al. (2019), which observed consistent improvements in maize and soybean (yield predicting) models in Canada when crop-specific masks were applied to spatially aggregate NDVI values. They highlighted the

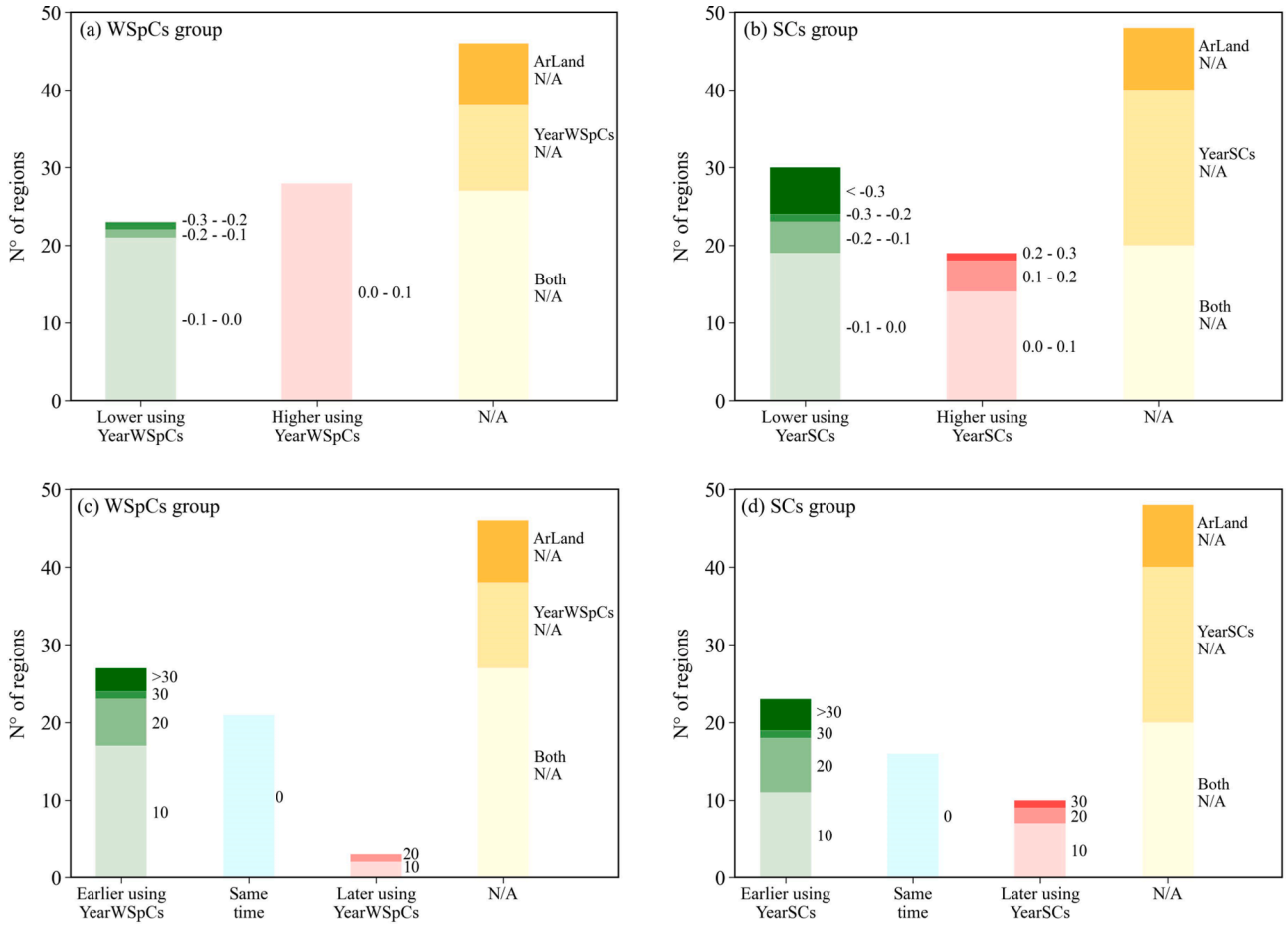


Fig. 12. Pairwise comparisons of the first-in-time significant RMSE: (a, c) YearWSpCs vs ArLand; (b, d) YearSCs vs ArLand. (a, b) RMSE residuals [t/ha]; (c, d) time-occurrence residuals [days].

importance of the exploitation of annual crop-specific masks in a DSS, being the more appropriate to select remote sensing pixels relevant for crop yield forecasts.

Results represent a methodological follow-up in crop mask exploitation compared to Baruth and Kucera (2006), where crop masks were built thanks to a disaggregation approach relying on land cover masks and statistical records, under the basic assumption that crops are spatial equally distributed within an administrative region.

With this work we contributed to confirm that NDVI time series increased their crop yield predictivity in reason of an increased specificity. This was particularly clear when spatial aggregations were performed using annual crop group-specific masks. Focusing on regional temporal NDVI profiles, we observed a progressive finer delineation of profiles at regional level in the transitions from spatial aggregations retrieved with generic arable land mask, to historical crop group-specific mask and finally, annual based crop group-specific masks.

In regions characterized by an even repartition of winter/spring and summer crop groups, historical and annual based crop group-specific masks allowed a better delineation of the NDVI time series compared to those derived by the generic arable land mask, facilitating the identification of two distinct temporal signatures of crop biomass accumulation within the agronomic season. Conversely, in regions where uneven proportions of winter/spring and summer crop groups prevailed, NDVI time series aggregated by means of a generic arable land mask remained close to that obtained from the prevalent crop group-specific masks, therefore, the correlations with yields resulted similar. Here, the main observed added value, was the greater ability in depicting the signal of the minority crop group and consequently the increased fitting when these data were correlated with reference yield values.

These findings, analysed in a context of operative DSS, allow a finer access on crop condition and crop biomass accumulation information, and are of great importance since they are provided at sub-national scale. For example, when NDVI data aggregated at regional level by means of crop group-specific masks are visually inspected against a reference temporal profile (e.g., a medium- or long-term trend NDVI profile), an easier and clearer interpretation of crop season's dynamics in terms of higher/lower biomass accumulation can be appreciated with respect to the reference profile. Furthermore, it allows a better recognition of crop's delays or advancements in the current growing season. These aspects are even more meaningful for those regions (i) characterized by an unbalanced share of winter/spring and summer crops, where it becomes more difficult to characterize the minority crop group and (ii) for regions characterized by a high share of rainfed crops, given that they are more exposed to inter-annual yield variability and abiotic stressors, for example drought.

4.2. *Crop masks comparison in yield forecasting*

Regarding the compared analyses on yield forecasting crop masks, we observed an overall correlation improvement in the results obtained by means of NDVI aggregated using annual crop group-specific masks, while a non-relevant effect was brought by the aggregation using historical crop group-specific masks. The obtained improvement was both numerical, retrieving better correlation results, and temporal, with significant correlation results occurring earlier in the agronomic seasons. These achievements, once implemented into a crop yield forecasting system, enable to retrieve more accurate forecasts and earlier-in-time yield previsions. Our findings confirmed previous contributions of

Liu et al. (2019), Zhang et al. (2019) and Shao et al. (2015) for case studies in Canada and Midwestern United States, respectively. These authors observed an increase in R^2 values ranging from +0.05 to +0.2 when crop-specific masks (even static) instead of generic masks were applied to spatially aggregate RS variables before feeding linear and/or non-linear regression yield predicting models. In our study, similar growths in R^2 values were achieved at a finer (regional) scale Europe-wide. Moreover, we proved that the introduction of annual crop group-specific masks (i.e., dynamic) provided an added value in terms of prediction timeliness and lead also to reduce forecasting errors for both winter/spring and summer crop groups. In some regions up to more than 0.3 t/ha (SCs group). The better correlation performances of NDVI aggregated using annual crop group-specific masks, were due to the capability in capturing inter-annual variations of crop-groups biomass productions and crop growth timing, especially at sub-national level. Hence, returning small changes in NDVI values (Gitelson, 2004). The increased skills of crop yield prediction observed by using annual masks were consistent with previous works (Mkhabela et al., 2011; Panek and Gozdoski, 2020), stating that a variation of NDVI by 0.1 units could result in an increase in grain yields from 0.2 to 2 t/ha, depending on the regions and crops of interest.

Looking at the results obtained for the two crop groups, larger benefits were observed for the SCs group than the WSpCs. More specifically, in SCs, annual crop group-specific masks led to a higher number of regions with significant and high correlation models, as well as a more marked rise in R^2 values (and decrease in RMSE) with respect to arable land and historical masks. One explanation could be that summer crops in the prevailing agricultural areas of Europe are generally more spatially clustered than winter/spring crops (Zhang et al., 2019), especially in Central Europe (Panek and Gozdoski, 2020), where the best results were retrieved in terms of R^2 values and RMSE. Furthermore, within SCs group, the yields of individual crops are quite correlated each other (Liu et al., 2019). This results in an advantage for yield forecasting, without the need to differentiate between single crops in the same crop group, even when the crops are mixed, both in time and space, in the same area. The quality of NDVI profiles could be another influencing factor: MODIS NDVI time series in Europe are more cloud-contaminated during the winter/spring season than during the summer crop season, as already proven by many authors (Armitage et al., 2013; King et al., 2013; Whitcraft et al., 2015; Slawik et al., 2016). However, it should be noticed that even for WSpCs group we identified some improvements in yield prediction when introducing annual crop group-specific indicators, although to a quantitative lesser extent compared to SCs. For WSpCs group, annual crop group-specific masks could help in returning satisfying R^2 values and significant RMSE from 10 to more than 30 days in advance with respect to static masks, providing an improvement in the timeliness of yield estimates. Retrieving proper annual masks for SCs group could be challenging in some cases, but whenever available, annual crop group-specific indicators led to higher R^2 values (+0.2) and lower RMSE (<-0.1 t/ha) quite early in the season, providing an improvement both in the accuracy and timeliness of yield estimates. These improvements brought by the use of annual masks could also be extended to the two most representative crops within each crop group, namely soft wheat and grain maize. This gives greater interest in exploiting RS annual crop group-specific indicators for yield forecasting, considering the importance of these two crops for Europe (Kelly, 2019; Svanidze and Đurić, 2021; Revilla et al., 2021).

The advantages obtained by means of the annual masks were widespread throughout the considered study area, but with regional concerns. Improvements were not highlighted for regions where (at least) one of the following five conditions were found. (i) Summer crops were absent or poorly cultivated; (ii) agricultural areas were large and/or homogeneously spread; (iii) the available reference time series were too short for computing a reliable regression model (six are the minimum years according to Nagy et el., 2018); (iv) the shares of cultivated areas were low, likely to non-relevant agricultural regions. To cope with this latter case, Liu et al. (2019) suggested to introduce new variables into the models, for example meteorological data or to increase the spatial resolution of RS data. Both these approaches were studied to estimate wheat yields in Latvia and Belgium, with limited improvements (Vannoppen et al., 2020; Vannoppen and Gobin, 2021). Finally, (v) introducing annual crop group-specific masks was challenging in regions where crops distributions within the same group were highly variable in time and it was not possible to define a prevalent crop. The greatest difficulties in this sense were observed in regions where the proportion between crops progressively changed during the reference time period, as observed from the regional statistics. Therefore, the annual masks were marked by crops shift, while static masks were unable to catch these changes. In the case of Europe, this last hypothesis is confirmed by previous studies, which reported a transition between rapeseed and wheat cultivations (Rondanini et al., 2012; Eurostat website).

Despite the potential of the implementation of annual crop group-specific masks into an operative yield forecasting system, it should also be highlighted that a limiting factor in their application is the availability of crop masks for the ongoing season (Zhang et al., 2019), because not ready in time for yield forecasting. With reference to this, there are two possible approaches individuated as potential solutions and deserving further research investigations: the first one relates to the use of the "previous year mask", which was highlighted as a good proxy for the distribution of crops in the current year (Becker-Reshef et al., 2018); the second one refers to the design and implementation of an algorithm for identifying in-season crop group-specific pure pixels and exploiting the last available annual crop group-specific indicators as a reference database.

5. Conclusions

This study proposed a compared analysis between crop masks which serve to spatially aggregate Remote Sensing indicators and improve the crop yield predictability in Europe at regional-scale level. Generic arable land mask, crop group-specific static masks and annual crop group-specific masks were tested. Results proved that the use of annual masks brought benefits for both crops monitoring and yield estimation at regional scale in EU. Specifically, annual crop group-specific masks favoured better analysis and interpretation of NDVI time series and improved accuracy and timeliness of crop yield forecasts. General improvements were observed in the use of annual crop group-specific masks for data aggregation in the whole study area. The best performances for annual masks were found for regions characterized by homogeneous shares of winter/spring and summer crops areas, and for soft wheat and grain maize areas, mostly in Central and Eastern Europe. Regarding the two crop categories, for WSpCs group we pointed out major advantages in the earliness of the estimate but limited improvements in the accuracy, while for SCs group both the accuracy and the timeliness of the estimate significantly improved. In order to make operational use of annual masks for a yield forecasting system, an issue still to be addressed remains the availability of in-season crop group-specific masks.

Funding

This work was supported by AGRI4CAST project [Project Portfolio 20,151 - Agri-CC-Resilience PRJ Id 30488], specifically deliverable D202205 – R&D in crop monitoring and modelling. The funding source had no involvement in the study design; in the collection, analysis and interpretation of the data; in the writing of the paper and in the decision to submit the article for publication.

CRediT authorship contribution statement

Giulia Ronchetti: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft,

Writing – review & editing, Visualization. **Giacinto Manfron:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Supervision. **Christof J. Weissteiner:** Writing – review & editing, Supervision. **Lorenzo Seguini:** Writing – review & editing, Supervision. **Luigi Nisini Scacchiafichi:** Data curation, Writing – review & editing. **Lorenzo Panarello:** Software, Writing – review & editing. **Bettina Baruth:** Writing – review & editing, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank all whose comments improved the paper. The authors also thank the AGRI4CAST project of the Joint Research Centre which financing supported the research activities included in this paper.

References

Ansarifar, J., Wang, L., Archontoulis, S.V., 2021. An interaction regression model for crop yield prediction. Sci. Rep. 11 (1), 1–14.
Armitage, R.P., Alberto Ramirez, F., Mark Danson, F., Ogunbadewa, E.Y., 2013. Probability of cloud-free observation conditions across Great Britain estimated using MODIS cloud mask. Remote Sens. Lett. 4 (5), 427–435.
Atzberger, C., 2013. Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. Remote Sens. (Basel) 5 (2), 949–981.
Bandyopadhyay, K.K., Pradhan, S., Sahoo, R.N., Singh, R., Gupta, V.K., Joshi, D.K., Sutradhar, A.K., 2014. Characterization of water stress and prediction of yield of wheat using spectral indices under varied water and nitrogen management practices. Agric. Water Manag. 146, 115–123.
Baruth, B., Kucera, L., 2006. Crop masking—Needs for the MARS crop yield forecasting system. Proceedings of the ISPRS 36 (8), W48.
Basso, B., Cammarano, D., Carfagna, E., 2013. Review of crop yield forecasting methods and early warning systems. In: Proceedings of the first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics, 241. FAO Headquarters, Rome, Italy.
Basso, B., Liu, L., 2019. Seasonal crop yield forecast: Methods, applications, and accuracies. Adv. Agron. 154, 201–255.
Becker-Reshef, I., Franch, B., Barker, B., Murphy, E., Santamaria-Artigas, A., Humber, M., Vermote, E., 2018. Prior season crop type masks for winter wheat yield forecasting: a US case study. Remote Sens. (Basel) 10 (10), 1659.
Becker-Reshef, I., Justice, C., Barker, B., Humber, M., Rembold, F., Bonifacio, R., Verdin, J., 2020. Strengthening agricultural decisions in countries at risk of food insecurity: the GEOGLAM crop monitor for early warning. Remote Sens. Environ. 237, 111553.
Boschetti, M., Busetto, L., Manfron, G., Laborte, A., Asilo, S., Pazhanivelan, S., Nelson, A., 2017. PhenoRice: a method for automatic extraction of spatio-temporal information on rice crops using satellite data time series. Remote Sens. Environ. 194, 347–365.
Boschetti, L., Flasse, S.P., Brivio, P.A., 2004. Analysis of the conflict between omission and commission in low spatial resolution dichotomic thematic products: the Pareto Boundary. Remote Sens. Environ. 91 (3–4), 280–292.
Chipanshi, A., Zhang, Y., Kouadio, L., Newlands, N., Davidson, A., Hill, H., Reichert, G., 2015. Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural landscape. Agric. For. Meteorol. 206, 137–150.
Cooper, M., Tang, T., Gho, C., Hart, T., Hammer, G., Messina, C., 2020. Integrating genetic gain and gap analysis to predict improvements in crop productivity. Crop Sci. 60 (2), 582–604.
Davidson, A.M., Fisette, T., McNairn, H., Daneshfar, B., Delince, J., 2017. Detailed crop mapping using remote sensing data (crop data layers). Handbk. Remote Sens. Agric. Statist. 91–117.
Eurostat, 2018. Regions in the European Union. Nomenclature of Territorial Units for Statistics. NUTS 2016/EU-28.
Eurostat website. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Main_annual_crop_statistics#Oilseeds. (Accessed 24 February 2022).
Fritz, S., See, L., Bayas, J.C.L., Waldner, F., Jacques, D., Becker-Reshef, I., McCallum, I., 2019. A comparison of global agricultural monitoring systems and current gaps. Agr. Syst. 168, 258–272.
Gitelson, A.A., 2004. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. J. Plant Physiol. 161 (2), 165–173.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.
Hatfield, J.L., Walthall, C.L., 2015. Meeting global food needs: realizing the potential via genetics× environment× management interactions. Agron. J. 107 (4), 1215–1226.
Hipólito de Sousa, J., Boscolo, D., Felipe Viana, B., 2018. Landscape and crop management strategies to conserve pollination services and increase yields in tropical coffee farms. Agr. Ecosyst. Environ. 256, 218–225.
Kelly, P. (2019). The EU cereals sector: Main features, challenges and prospects.
Kendall, M.G., 1975. Rank Correlation Methods, 4th edition. Charles Griffin, London.
King, M.D., Platnick, S., Menzel, W.P., Ackerman, S.A., Hubanks, P.A., 2013. Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites. IEEE Trans. Geosci. Remote Sens. 51 (7), 3826–3852.
Kosztra, B., Büttner, G., Hazeu, G., Arnold, S., 2017. Updated CLC illustrated nomenclature guidelines. Wien, Austria, European Environment Agency, pp. 1–124.
Kouadio, L., Newlands, N.K., Davidson, A., Zhang, Y., Chipanshi, A., 2014. Assessing the performance of MODIS NDVI and EVI for seasonal crop yield forecasting at the ecodistrict scale. Remote Sens. (Basel) 6 (10), 10193–10214.
Liu, J., Shang, J., Qian, B., Huffman, T., Zhang, Y., Dong, T., Martin, T., 2019. Crop yield estimation using time-series MODIS data and the effects of cropland masks in Ontario, Canada. Remote Sensing 11 (20), 2019.
Liu, J., Huffman, T., Qian, B., Shang, J., Li, Q., Dong, T., Jing, Q., 2020. Crop yield estimation in the Canadian Prairies using Terra/MODIS-derived crop metrics. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 2685–2697.
López-Lozano, R., Baruth, B., El Aydam, M., Willems, E., 2015a. & García Azcárate, T. Crop monitoring and yield forecasting at global level, The GLOBCAST project from the European Commission.
López-Lozano, R., Duveiller, G., Seguini, L., Meroni, M., García-Condado, S., Hooker, J., Baruth, B., 2015b. Towards regional grain yield forecasting with 1 km-resolution EO biophysical products: strengths and limitations at pan-European level. Agric. For. Meteorol. 206, 12–32.
Mann, H.B., 1945. Nonparametric tests against trend. Econometrica 245–259.
Meroni, M., Waldner, F., Seguini, L., Kerdiles, H., & Rembold, F., 2021. Yield forecasting with machine learning and small data: what gains for grains? *arXiv preprint arXiv: 2104.13246*.
Mkhabela, M.S., Bullock, P., Raj, S., Wang, S., Yang, Y., 2011. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. Agric. For. Meteorol. 151 (3), 385–393.
Nagy, A., Fehér, J., Tamás, J., 2018. Wheat and maize yield forecasting for the Tisza River catchment using MODIS NDVI time series and reported crop statistics. Comput. Electron. Agric. 151, 41–49.
National Agricultural Statistics Service (NASS), 2006. The yield forecasting program of NASS by the Statistical Methods Branch, Estimates Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C., May 2006. NASS Staff Report No. SMB 06-01.
Panek, E., Gozdowski, D., 2020. Analysis of relationship between cereal yield and NDVI for selected regions of Central Europe based on MODIS satellite data. Remote Sens. Appl.: Soc. Environ. 17, 100286.
Pérez-Hoyos, A., Udías, A., Rembold, F., 2020. Integrating multiple land cover maps through a multi-criteria analysis to improve agricultural monitoring in Africa. Int. J. Appl. Earth Obs. Geoinf. 88, 102064.
Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016. Sensitivity analysis of environmental models: a systematic review with practical workflow. Environ. Model. Softw. 79, 214–232.
Revilla, P., Alves, M.L., Andelković, V., Balconi, C., Dinis, I., Mendes-Moreira, P., Malvar, R.A., 2021. Traditional foods from maize (Zea mays L.) in Europe. Front. Nutr. 8.
Rojas, O., Vrieling, A., Rembold, F., 2011. Assessing drought probability for agricultural areas in Africa with coarse resolution remote sensing imagery. Remote Sens. Environ. 115 (2), 343–352.
Ronchetti, G., Weissteiner, C., Manfron, G., Panarello, L., Seguini, L., 2021. Crop group-specific probability masks: model description. European Commission JRC127408.
Ronchetti, G., Nisini Scacchiafichi, L., Seguini, L., Cerrani, I., Van der Velde, M., 2022. Consolidating and harmonising subnational crop statistics of the European Union. In Preparation With Earth System Science Data Discussions.
Rondanini, D.P., Gomez, N.V., Agosti, M.B., Miralles, D.J., 2012. Global trends of rapeseed grain yield stability and rapeseed-to-wheat yield ratio in the last four decades. Eur. J. Agron. 37 (1), 56–65.
Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., Harlan, J.C., 1974. Monitoring the Vernal Advancement of Retrogradation of Natural Vegetation; NASA/GSFC: Greenbelt. MD, USA.
Shahhosseini, M., Hu, G., Archontoulis, S.V., 2020. Forecasting corn yield with machine learning ensembles. Front. Plant Sci. 11, 1120.
Shammi, S.A., Meng, Q., 2021. Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling. Ecol. Ind. 121, 107124.
Shao, Y., Campbell, J.B., Taff, G.N., Zheng, B., 2015. An analysis of cropland mask choice and ancillary data for annual corn yield forecasting using MODIS data. Int. J. Appl. Earth Obs. Geoinf. 38, 78–87.
Skakun, S., Vermote, E., Roger, J.C., Franch, B., 2017. Combined use of Landsat-8 and Sentinel-2A images for winter crop mapping and winter wheat yield assessment at regional scale. AIMS Geosciences 3 (2), 163.

Slawik, A.J., Adelizzi, E.A., Koretsky, G.M., 2016. Assessment of Cloud Occurrence and Impact on Remote Sensing using Global. INSTITUTE FOR DEFENSE ANALYSES ALEXANDRIA VA, Open-Source, Scientific Datasets (REDACTED).
Svanidze, M., Đurić, I., 2021. Global wheat market dynamics: What is the role of the EU and the Black Sea wheat exporters? Agriculture 11 (8), 799.
Van der Velde, M., Van Diepen, C., Baruth, B., 2019. The European crop monitoring and yield forecasting system: celebrating 25 years of MARS bulletins. Agr. Syst. 168, 56–57.
Van Rossum, G., Drake, F.L., 2009. Python 3 Reference Manual. CreateSpace, Scotts Valley, CA.
Vannoppen, A., Gobin, A., Kotova, L., Top, S., De Cruz, L., Vīksna, A., Termonia, P., 2020. Wheat yield estimation from NDVI and regional climate models in Latvia. Remote Sens. (Basel) 12 (14), 2206.
Vannoppen, A., Gobin, A., 2021. Estimating farm wheat yields from NDVI and meteorological data. Agronomy 11 (5), 946.
Vermote, E.F., Roger, J.C., Ray, J.P., 2015. MODIS Surface Reflectance User's Guide (Collection 6). MODIS Land Surface Reflectance Science Computing Facility, Version 1 (4).
Waldner, F., Fritz, S., Di Gregorio, A., Defourny, P., 2015. Mapping priorities to focus cropland mapping activities: fitness assessment of existing global, regional and national cropland maps. Remote Sens. (Basel) 7 (6), 7959–7986.
Weissteiner, C.J., López-Lozano, R., Manfron, G., Duveiller, G., Hooker, J., Van der Velde, M., Baruth, B., 2019. A crop group-specific pure pixel time series for Europe. Remote Sens. (Basel) 11 (22), 2668.
Whitcraft, A.K., Vermote, E.F., Becker-Reshef, I., Justice, C.O., 2015. Cloud cover throughout the agricultural growing season: impacts on passive optical earth observations. Remote Sens. Environ. 156, 438–447.
Wu, Z., Huang, N.E., Long, S.R., Peng, C.K., 2007. On the trend, detrending, and variability of nonlinear and nonstationary time series. Proc. Natl. Acad. Sci. 104 (38), 14889–14894.
Zhang, Y., Chipanshi, A., Daneshfar, B., Koiter, L., Champagne, C., Davidson, A., Bédard, F., 2019. Effect of using crop specific masks on Earth Observation based crop yield forecasting across Canada. Remote Sens. Appl.: Soc. Environ. 13, 121–137.
Ziliani, M.G., Altaf, M.U., Aragon, B., Houborg, R., Franz, T.E., Lu, Y., McCabe, M.F., 2022. Early season prediction of within-field crop yield variability by assimilating CubeSat data into a crop model. Agric. For. Meteorol. 313, 108736.